Columbia University Automatic Video Search in TRECVID 2006

DTO
Video Analysis Content Extraction

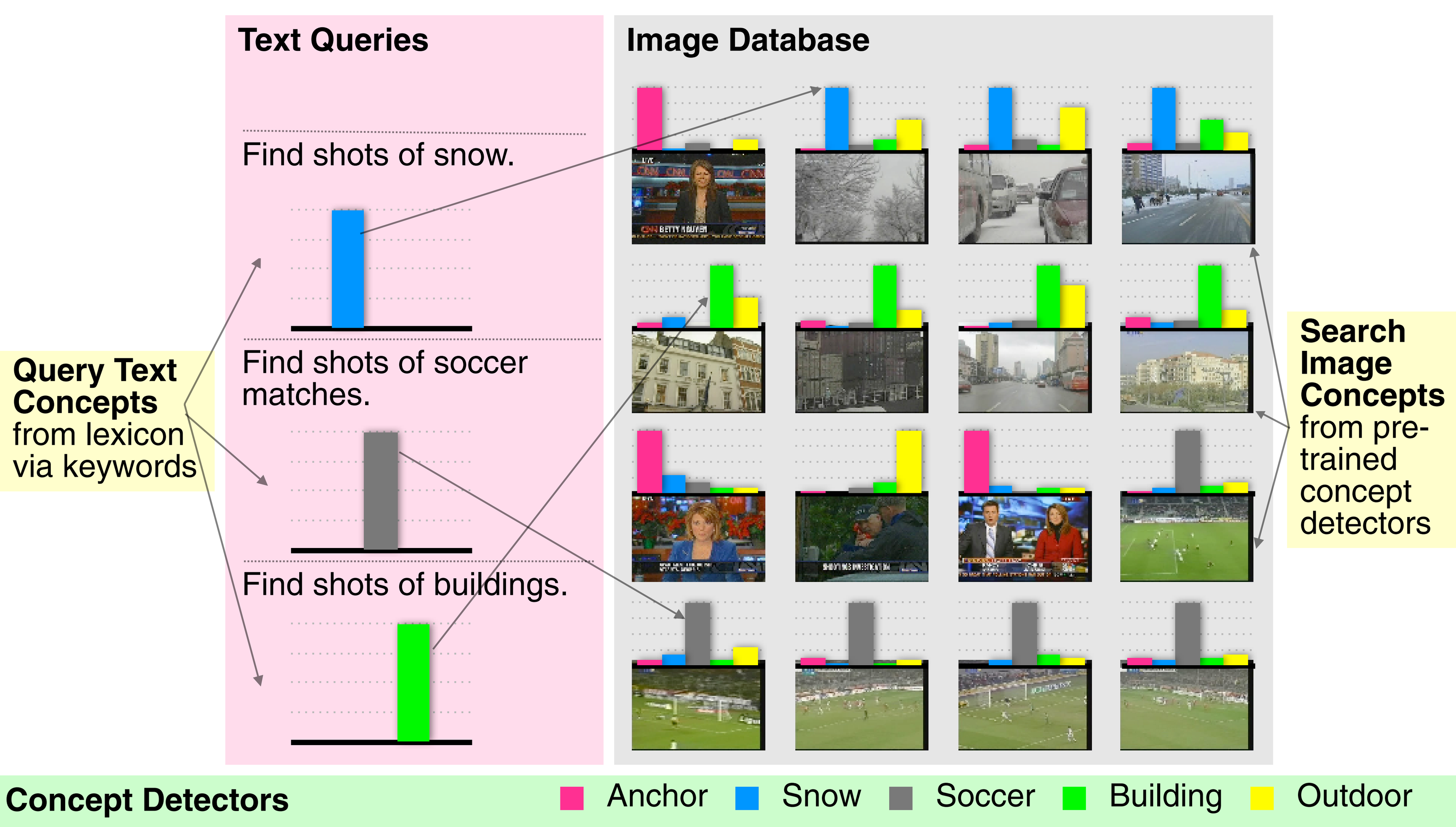
Shih-Fu Chang, Winston Hsu, Wei Jiang, Lyndon Kennedy,
Dong Xu, Akira Yanagawa, and Eric Zavesky

Objectives and Unique Features

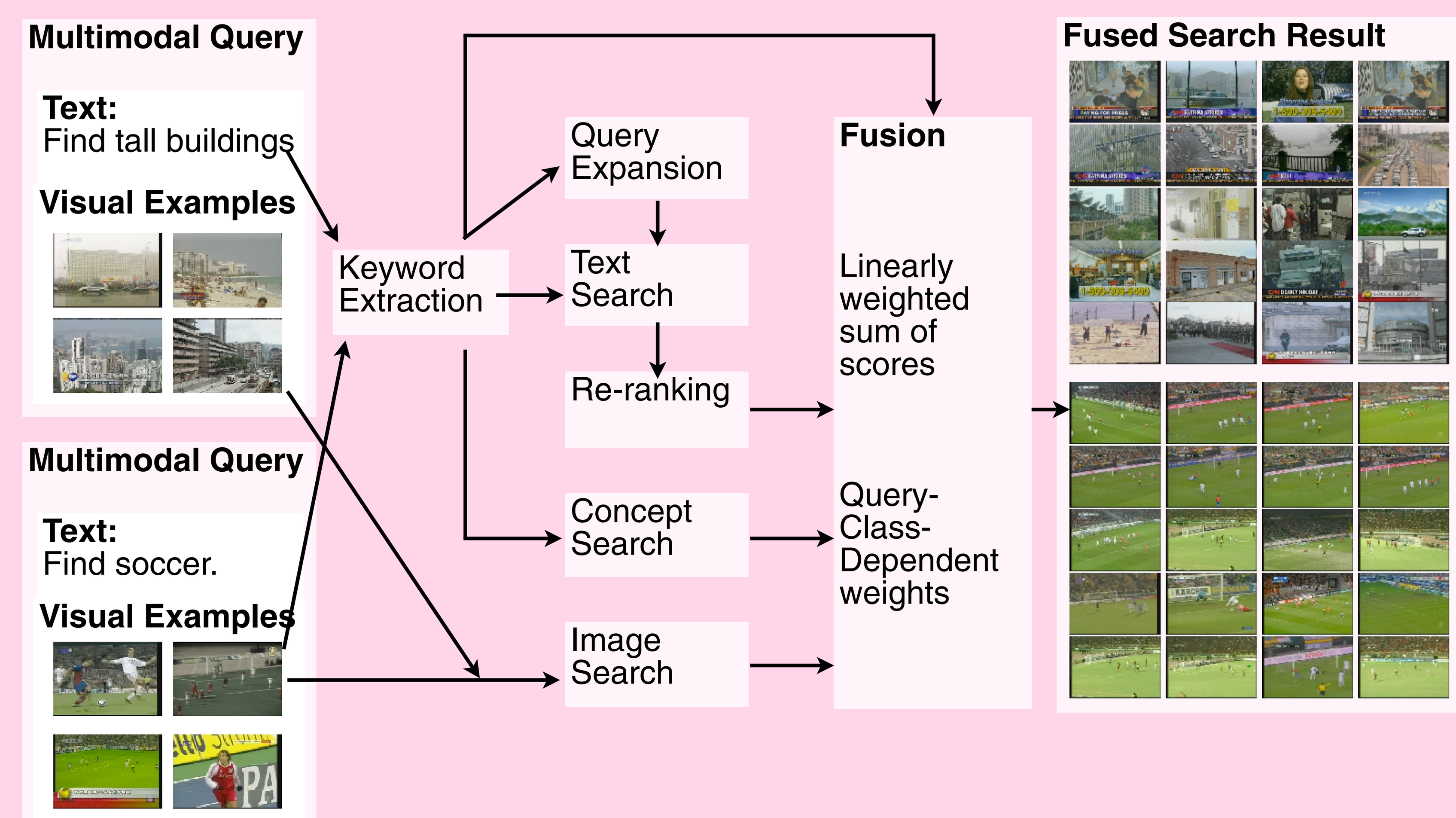
+ **Objective:** Enable semantic search over a large video collection. Leverage visual content and text.
+ Test *Concept-Based Search* using a large set of pre-trained concept detectors. (374 visual concepts).
+ Supplement concept-based search with other methods through *Query-Class-Dependent Models.*
+ Analyze impact of fusion of multimodal search methods and query-class-dependency in fully automatic search.

Concept-Based Search

+ Text-based queries against visual content.
+ Concept detectors predict the concept in image.
+ Lexicon of definitions maps keyword to concept.



Search System Overview



Text Search: keywords against story-separated speech recognition text.
Re-ranking: discover visually recurrent patterns using information bottleneck.
Query Expansion: mine names from text contemporary to search set.
Image Search: query-by image example, using distance and SVMs.
Concept Search: match text keywords to pre-trained concept detectors.

Large-Scale Concept Lexicon

+ Leverage LSCOM annotations. (449 concepts).
+ 374 detectable concepts, compared to standard 39.
+ 10x increase in lexicon – 30% increase in search

	39 Concepts	374 Concepts	Effect
Search Performance[1]	TV2005 - MAP: **.035** TV2006 - MAP: **.019**	TV2005 - MAP: **.078** TV2006 - MAP: **.025**	**30-100%** increase in retrieval performance
Query Coverage	**11-12** Query Matches **1.1-1.8** Concepts/Query	**17** Query Matches **1.3-2.5** Concepts/Query	Smaller than **2x** increases in coverage
Concept Frequency	Average: **5000** examples	Average: **1200** examples	**75%** decrease in freq. Min(39) > Median(374)
Detection Quality[2]	MAP: **0.39** (over 2005 validation data)	MAP: **0.26** (over 2005 validation data)	**33%** decrease in concept detection accuracy

[1] Using Columbia Concept-Based Search Method (Described above). [2] Using Columbia Baseline Concept Detection.

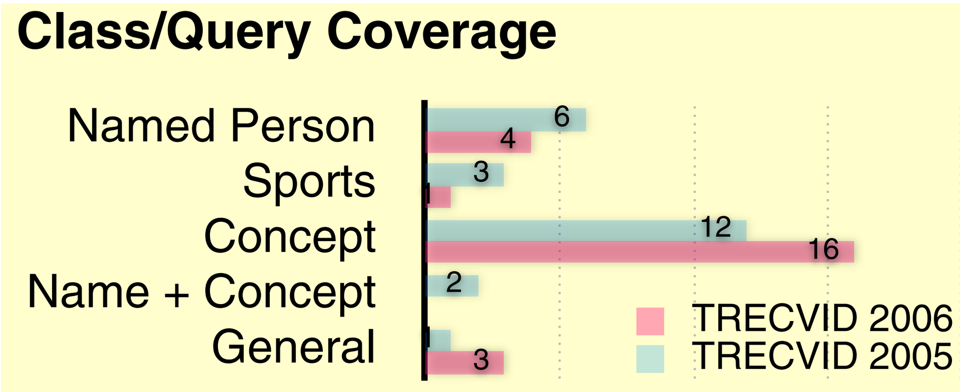
Query-Class-Dependent Models

Named Person: if named entity detected in query. Rely on text search.
Sports: if sports keyword detected in query. Rely on visual examples.
Concept: if keyword maps to concept detector. Rely on concept search.
Named Person + Concept: if both named entity and concept detected. Combine text and concept search equally.
General: for all other queries. Combine text and visual examples equally.

Class Preparation

+ Select classes to suit tools.
+ Weights optimize AP (over training queries).
+ Classify by query keywords. (automatically at query time)

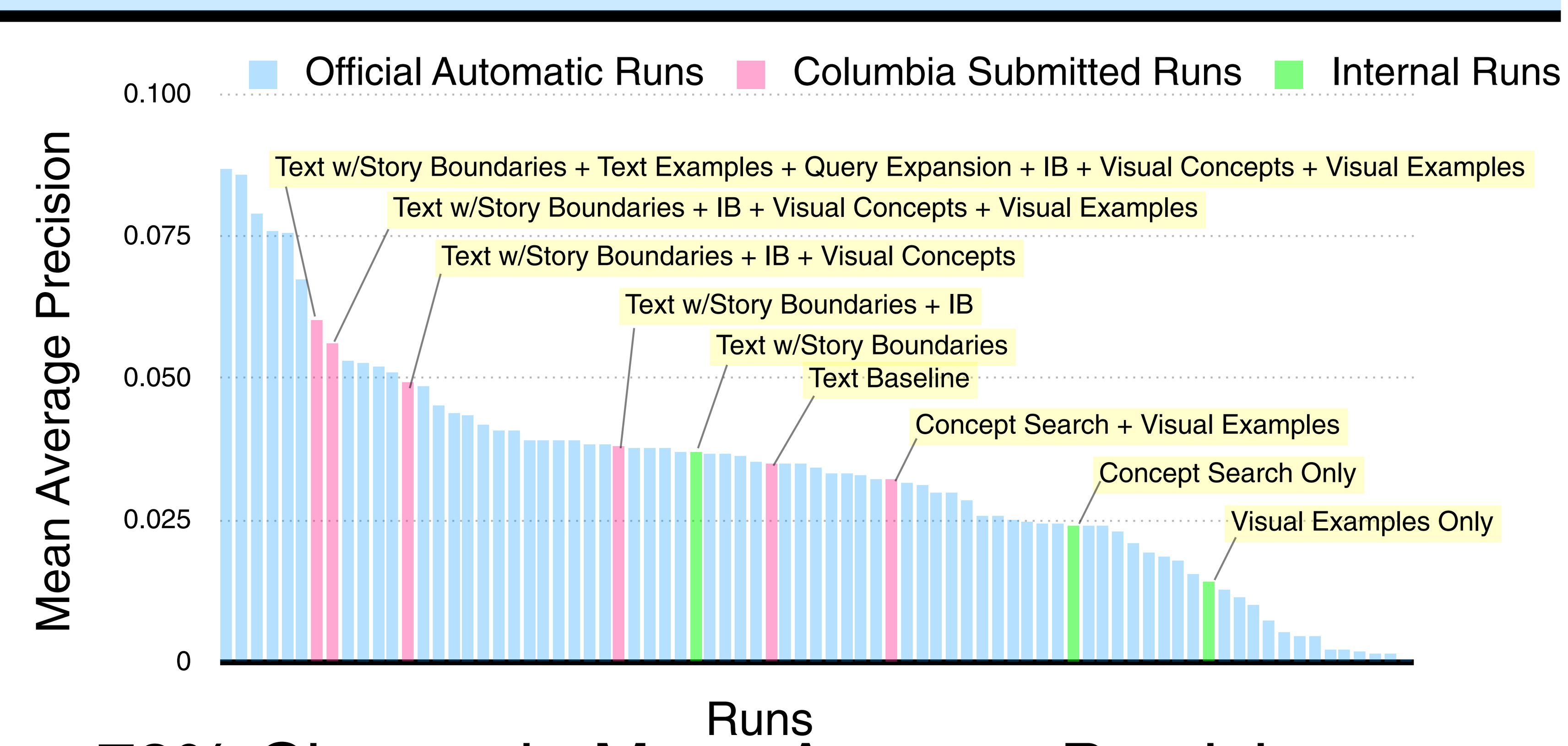
Class/Query Coverage



Evaluation

	Text	IB	Concept	Visual	Fused	Change
Building	0.04	0.03	0.30	0.10	0.33	**+10%**
Condi Rice	0.19	0.20	0.00	0.01	0.20	**+0%**
Soccer	0.34	0.42	0.58	0.42	0.83	**+43%**
Snow	0.19	0.29	0.33	0.03	0.80	**+143%**
All (2006)	**0.08**	**0.09**	**0.10**	**0.07**	**0.19**	**+90%**

+ Tested over 160+ hours of multilingual news video: TRECVID 2006.
+ Consistent gains in P@100 (Precision of top 100) through class-dependent multimodal fusion.
+ Text and Concept-based searches perform well.
+ Most query topics have matching concepts.



+ **70%** Change in Mean Average Precision (MAP), using multimodal search.
+ **30%** Change in MAP from concept-based search

LSCOM Lexicon and Annotations
http://www.ee.columbia.edu/dvmm/lscom

Digital Video and Multimedia Lab
http://www.ee.columbia.edu/dvmm

CuVid Video Search Engine
http://www.ee.columbia.edu/cuvidsearch